NORTH-HOLLAND

# Monotone Gram Matrices and Deepest Surrogate Inequalities in Accelerated Relaxation Methods for Convex Feasibility Problems

Krzysztof C. Kiwiel
*Systems Research Institute*
*Newelska 6*
*01-477 Warsaw, Poland*
*Email: hiwiel@ibspan.waw.pl*

Submitted by Richard A. Brualdi

ABSTRACT

The relaxation method for linear inequalities iterates by projecting the current point onto the most violated constraint. Accelerated methods project onto the intersection of several halfspaces or onto a surrogate halfspace corresponding to a nonnegative combination of constraints. We extend Todd's conditions for finding best surrogate inequalities via the solution of systems of linear equations. Our techniques may be used for accelerating various methods for convex feasibility and optimization problems. © *Elsevier Science Inc., 1997*

## 1. INTRODUCTION

The relaxation method [Agm54, MoS54] for linear inequalities iterates by projecting the current point onto the halfspace corresponding to the most violated inequality. This method may suffer from slow convergence if the solution set is "flat" [Gof80], and an old remedy [Mer62] is to use surrogates (nonnegative combinations) of violated inequalities. The best surrogate [BGT81, GoT82] corresponds to projecting on the intersection of halfspaces of violated inequalities. This may, in general, be too expensive.

In this article we extend Todd's [Tod79] conditions for finding best surrogates via the solution of systems of linear equations. Our techniques may be used for accelerating various methods for set intersection, convex feasibil-

*LINEAR ALGEBRA AND ITS APPLICATIONS* 252:27–33 (1997)



0024-3795/97/$17.00
SSDI 0024-3795(95)00608-7

ity, and optimization problems; see, e.g., [BB96, Cen81, DPI85, GP93, Kiw95, Kiw96a, Kiw96b, Oko92, Ott88, YaM92] and references therein.

In Section 2, as in [GoT82], we highlight the connection between deepest surrogates and monotonicity of certain Gram matrices. Ways of ensuring this monotonicity are studied in Section 3.

We use the following notation. We denote by $\langle \cdot , \cdot \rangle$ and $|\cdot|$ respectively the usual inner product and norm in $\mathbb{R}^n$. $P_S(x) = \arg\min_{y \in S} |x - y|$ denotes the *projection* of $x$ onto a closed convex set $S$ in $\mathbb{R}^n$, and $d_S(x) = \inf_{y \in S} |x - y|$. For any set $\mathscr{A} \subset \mathbb{R}^n$, $\operatorname{lin} \mathscr{A}$ denotes its *linear span*, $\operatorname{cone} \mathscr{A} = \{a: a = \sum_{i=1}^m \lambda_i a^i, a^i \in \mathscr{A}, \lambda_i \geqslant 0, m < \infty\}$ denotes its convex *conical hull*, $\mathscr{A}^+ = \{x: \langle x, y \rangle \geqslant 0 \, \forall y \in \mathscr{A}\}$ denotes its *positive polar cone*, and $\mathscr{A}^\perp = \{x: \langle x, y \rangle = 0 \, \forall y \in \mathscr{A}\}$ denotes its *orthogonal subspace*. For a matrix $A \in \mathbb{R}^{n \times m}$, $a_{ij}$ and $a^i$ denote its $ij$th element and $i$th column respectively. Given a set $\mathscr{I} \subset \{1:m\}$, $A_{\mathscr{I}}$ denotes the matrix with columns $a^i, i \in \mathscr{I}$. Matrix inequalities hold componentwisely. We let $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n: x \geqslant 0\}$ and $t_+ = \max\{t, 0\} \, \forall t \in \mathbb{R}$. We need the following well-known result.

LEMMA 1.1. *A matrix $G \in \mathbb{R}^{m \times m}$ is monotone, i.e., $\{\nu: G\nu \geqslant 0\} \subset \mathbb{R}^m_+$, iff $G^{-1} \geqslant 0$.*

*Proof.* Suppose $G$ is monotone. If $G\nu = 0$ then $\nu \geqslant 0$, and $\nu \leqslant 0$ from $G(-\nu) = 0$, so $\nu = 0$ and $G^{-1}$ exists. For any $\nu \geqslant 0, 0 \leqslant \nu = G(G^{-1}\nu)$ yields $G^{-1}\nu \geqslant 0$, so $G^{-1} \geqslant 0$. Conversely, if $G^{-1} \geqslant 0$ and $G\nu \geqslant 0$ then $\nu = G^{-1}G\nu \geqslant 0$. ∎

## 2. DEEPEST SURROGATE INEQUALITIES

Given $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^m$, consider the system of linear inequalities $\langle a^i, x \rangle \leqslant b_i, i = 1:m$, having a (possibly empty) solution set $\mathscr{P} = \{x: A^T x \leqslant b\}$. Suppose $a^i \neq 0$ for $i = 1:m$. Then each inequality defines a closed halfspace $H_i = \{x: \langle a^i, x \rangle \leqslant b_i\}$. In the classical relaxation method for finding a point in $\mathscr{P} = \bigcap_{i=1}^m H_i$, given a current point $\tilde{x} \notin \mathscr{P}$, one finds the next point $\bar{x}$ by projecting $\tilde{x}$ on the halfspace $H_i$ that is furthest from $\tilde{x}$, since for faster convergence one wishes to maximize $|\bar{x} - \tilde{x}|$. By combining inequalities one can sometimes obtain halfspaces that are further from $\tilde{x}$.

If $\lambda \in \mathbb{R}^m_+, a^\lambda = A\lambda$, and $b_\lambda = b^T\lambda$, then the *surrogate inequality* $\langle a^\lambda, x \rangle \leqslant b_\lambda$ is valid ($A^T x \leqslant b \Rightarrow \lambda^T A^T x \leqslant \lambda^T b$). The *deepest* surrogate inequality that maximizes the distance $(\langle a^\lambda, \tilde{x} \rangle - b_\lambda)_+ / |a^\lambda|$ from $\tilde{x}$ to $H_\lambda = \{x: \langle a^\lambda, x \rangle \leqslant b_\lambda\}$ corresponds to

$$\tilde{\lambda} \in \operatorname{Arg\,max}\{\tilde{s}^T\lambda / |A\lambda|: \lambda \geqslant 0\}, \tag{2.1}$$

where $\tilde{s} := A^T\tilde{x} - b \nleqslant 0$ ($\tilde{x} \notin \mathscr{P}$). Clearly, if $\mathscr{P} \neq \varnothing$ then $H_{\tilde{\lambda}}$ is the unique halfspace containing $\mathscr{P}$ that is furthest from $\tilde{x}$, and $H_{\tilde{\lambda}} = \{x: \langle \tilde{d}, x - \tilde{x}\rangle \geqslant |\tilde{d}|^2\}$, where $\tilde{d} = P_{\mathscr{P}}(\tilde{x}) - \tilde{x}$ (since for any halfspace $H \supset P_{\mathscr{P}}(\tilde{x})$, $d_H(\tilde{x}) < d_{\mathscr{P}}(\tilde{x})$ unless $P_H(\tilde{x}) = P_{\mathscr{P}}(\tilde{x})$). Of course, $\tilde{d}$ solves the quadratic programming (QP) problem

$$\tilde{d} = \arg\min\left\{|d|^2/2\colon A^Td \leqslant -\tilde{s}\right\}. \tag{2.2}$$

By duality, we may equivalently find its (possibly nonunique) Lagrange multiplier vector

$$\tilde{\lambda} \in \operatorname{Arg}\min\left\{|A\lambda|^2/2 - \tilde{s}^T\lambda\colon \lambda \geqslant 0\right\}. \tag{2.3}$$

Indeed, by the Karush–Kuhn–Tucker conditions, $\tilde{d}$ and $\tilde{\lambda}$ satisfy (2.2)–(2.3) iff $\tilde{d} = -A\tilde{\lambda}$, $A^T\tilde{d} \leqslant -\tilde{s}$, $\tilde{\lambda} \geqslant 0$, and $\tilde{\lambda}^T(A^T\tilde{d} + \tilde{s}) = 0$. Hence $\tilde{s}^T\tilde{\lambda} = |\tilde{d}|^2$ and $\langle a^{\tilde{\lambda}}, \tilde{x}\rangle - |\tilde{d}|^2 = \tilde{\lambda}^TA^T\tilde{x} - \tilde{x}^TA\tilde{\lambda} + b^T\tilde{\lambda} = b_{\tilde{\lambda}}$, so $\langle a^{\tilde{\lambda}}, x\rangle \leqslant b_{\tilde{\lambda}}$ iff $\langle \tilde{d}, x - \tilde{x}\rangle \geqslant |\tilde{d}|^2$, and $P_{H_{\tilde{\lambda}}}(\tilde{x}) = P_{\mathscr{P}}(\tilde{x})$.

Of course, finding the deepest surrogate inequality via (2.1)–(2.3) may be too expensive. The next result helps in selecting subsets of inequalities for which the projections are "easy." For any $\mathscr{J} \subset \{1{:}m\}$, let $\mathscr{A}_{\mathscr{J}} = \{a^i\}_{i\in\mathscr{J}}$, $\mathscr{P}_{\mathscr{J}} = \{x\colon A^T_{\mathscr{J}}x \leqslant b_{\mathscr{J}}\}$, $\mathscr{D}_{\mathscr{J}} = \{d\colon A^T_{\mathscr{J}}d \leqslant -\tilde{s}_{\mathscr{J}}\}$, and $G_{\mathscr{J}\mathscr{J}} = A^T_{\mathscr{J}}A_{\mathscr{J}}$. If $\mathscr{P}_{\mathscr{J}} \neq \varnothing$, let $\tilde{d}_{(\mathscr{J})} = \mathscr{P}_{\mathscr{D}_{\mathscr{J}}}(0)$, so that $\tilde{d}_{(\mathscr{J})} = P_{\mathscr{P}_{\mathscr{J}}}(\tilde{x}) - \tilde{x}$ from $\tilde{s}_{\mathscr{J}} = A^T_{\mathscr{J}}\tilde{x} - b_{\mathscr{J}}$.

LEMMA 2.1. *Suppose* $\mathscr{J} \subset \{1{:}m\}$ *and* rank $A_{\mathscr{J}} = |\mathscr{J}|$. *Let* $\bar{\lambda}_{\mathscr{J}} = G^{-1}_{\mathscr{J}\mathscr{J}}\tilde{s}_{\mathscr{J}}$ *and* $\bar{d}_{(\mathscr{J})} = -A_{\mathscr{J}}\bar{\lambda}_{\mathscr{J}}$. *Then* $\bar{d}_{(\mathscr{J})} = \arg\min\,\{|d|^2/2\colon A^T_{\mathscr{J}}d = -\tilde{s}_{\mathscr{J}}\}$. *Further,* $\bar{d}_{(\mathscr{J})} = \tilde{d}_{(\mathscr{J})} \Leftrightarrow \bar{\lambda}_{\mathscr{J}} \geqslant 0 \Leftrightarrow$

$$\bar{\lambda}_{\mathscr{J}} \in \operatorname{Arg}\max\left\{\tilde{s}^T_{\mathscr{J}}\lambda_{\mathscr{J}}/|A_{\mathscr{J}}\lambda_{\mathscr{J}}|\colon \lambda_{\mathscr{J}} \geqslant 0\right\}. \tag{2.4}$$

*In particular,* (2.4) *holds if* $\tilde{s}_{\mathscr{J}} \geqslant 0$ *and* $G^{-1}_{\mathscr{J}\mathscr{J}} \geqslant 0$.

*Proof.* The corresponding QP problems have unique solutions and multipliers. ■

The next section shows how to choose $\mathscr{J}$ with $G^{-1}_{\mathscr{J}\mathscr{J}} \geqslant 0$, as required in Lemma 2.1.

## 3. CHOOSING A MONOTONE GRAM MATRIX

Our first result on monotone Gram matrices facilitates geometric interpretations.

LEMMA 3.1. *Let $G = A^TA$ be the Gram matrix of $A \in \mathbb{R}^{n\times m}$ and let $\mathscr{A} = \{a^i\}_{i=1}^m$. Then $G$ is monotone iff* rank $A = m$ *and* $\mathscr{A}^+ \cap \operatorname{lin}\mathscr{A} \subset \operatorname{cone}\mathscr{A}$, *i.e., cone $\mathscr{A}$ is obtuse in* lin $\mathscr{A}$.

*Proof.* If $G$ is monotone and $A\mu \in \mathscr{A}^+$ for some $\mu$ then $G\mu = A^T A\mu \geqslant 0$ yields $\mu \geqslant 0$, so $A\mu \in \operatorname{cone}\mathscr{A}$, whereas rank $A$ = rank $G = m$ by Lemma 1.1. For the converse, if $A^TA\mu \geqslant 0$ and $A\mu = A\mu'$ for some $\mu' \geqslant 0$ then rank $A = m$ yields $\mu = \mu' \geqslant 0$. ∎

The first part of the following result is well known (cf. [Bjö90, DBMS79, LaH74]).

LEMMA 3.2. *Let $A \in \mathbb{R}^{n\times m}$ have* rank $A = m$, *let $G = A^TA$, let $A^\dagger = G^{-1}A^T$, let $\bar a \in \mathbb{R}^n$, let $R$ be an $m \times m$ upper triangular matrix satisfying $R^TR = G$, let $r, \bar\mu \in \mathbb{R}^m$ satisfy $R^Tr = A^T\bar a$ and $R\bar\mu = -r$, and let $\rho = (|\bar a|^2 - |r|^2)^{1/2}$. Then $\bar\mu = \arg\min_\mu |A\mu + \bar a| = -A^\dagger \bar a$ and $\rho = |A\bar\mu + \bar a|$. Moreover, if $\bar A = [A, \bar a]$, $\bar G = \bar A^T\bar A$ and $\bar R = \begin{bmatrix} R & r \\ 0 & \rho \end{bmatrix}$, then $\bar R^T\bar R = \bar G$ and* rank $\bar R$ = rank $\bar A$, *with* rank $\bar A = m + 1$ *iff $\rho \neq 0$. Further,*

(*i*) *$\rho \neq 0$ iff $\bar G^{-1}$ exists, in which case*

$$\bar G^{-1} = \begin{bmatrix} G^{-1} + \bar\mu\bar\mu^T/\rho^2 & \bar\mu/\rho^2 \\ \bar\mu^T/\rho^2 & 1/\rho^2 \end{bmatrix} = \bar R^{-1}\bar R^{-T},$$

$$\bar R^{-1} = \begin{bmatrix} R^{-1} & -R^{-1}r/\rho \\ 0 & 1/\rho \end{bmatrix} = \begin{bmatrix} R^{-1} & \bar\mu/\rho \\ 0 & 1/\rho \end{bmatrix}.$$

(*ii*) *$\bar G$ is monotone iff $\rho \neq 0$, $\bar\mu \geqslant 0$ and $\rho^2G^{-1} + \bar\mu\bar\mu^T \geqslant 0$.*

(*iii*) *$\bar\mu \geqslant 0$ iff $\bar a \in -\operatorname{cone}\mathscr{A} + \mathscr{A}^\perp$, where $\mathscr{A} = \{a^i\}_{i=1}^m$.*

(*iv*) *If $G$ is monotone and $\bar\mu \geqslant 0$ then $\bar G$ is monotone $\Leftrightarrow \rho > 0 \Leftrightarrow \exists \check x$: $\bar A^T\check x < 0 \Leftrightarrow \exists \check x$: $\bar a^T\check x < 0$ and $A^T\check x \leqslant 0 \Leftrightarrow \bar a \neq 0$ and $\exists \check x$: $\bar a^T\check x \leqslant 0$ and $A^T\check x < 0$.*

(*v*) *If $G^{-1} \geqslant 0$, $\rho > 0$, and $A^T\bar a \leqslant 0$ then $\bar\mu \geqslant 0$ and $\bar G^{-1} \geqslant 0$.*

(*vi*) *If $R^{-1} \geqslant 0$, $\rho > 0$, and $A^T\bar a \leqslant 0$ then $\bar R^{-1} \geqslant 0$ and $\bar G^{-1} \geqslant 0$.*

*Proof.* Since $|A\mu| = |R\mu| \forall\mu$, rank $A$ = rank $R$, $R^{-1}$ exists and $\overline{R}^T\overline{R} = \overline{A}^T\overline{A}$ iff

$$\overline{G} = \overline{A}^T\overline{A} = \begin{bmatrix} A^TA & A^T\overline{a} \\ \overline{a}^TA & |\overline{a}|^2 \end{bmatrix} = \begin{bmatrix} R^TR & R^Tr \\ r^TR & |r|^2 + \rho^2 \end{bmatrix} = \overline{R}^T\overline{R},$$

with $|A\mu + \overline{a}|^2 = |\overline{A}\begin{pmatrix}\mu \\ 1\end{pmatrix}|^2 = |R\mu + r|^2 + \rho^2$ minimized by $\overline{\mu} = -R^{-1}r$. Assertion (i) is verified by direct calculation. Assertion (ii) follows from (i) and Lemma 1.1. To prove (iii), note that $\overline{a} = P_{\text{lin}\,\mathscr{A}}\overline{a} + P_{\mathscr{A}^\perp}\overline{a}$, where $P_{\text{lin}\,\mathscr{A}}\overline{a} = -A\overline{\mu}$. As for (iv), if $\overline{G}^{-1} \geqslant 0$ (cf. Lemma 1.1), let $\check{x} = -\overline{A}\overline{G}^{-1}e$, where $e = (1,\dots,1)^T$, to get $\overline{A}^T\check{x} < 0$. Conversely, if $\overline{a}^T\check{x} < 0$ and $A^T\check{x} \leqslant 0$ then $\rho \neq 0$, since otherwise $\overline{a} = -A\overline{\mu}$ would yield $0 > \overline{a}^T\check{x} = -\overline{\mu}^TA^T\check{x} \geqslant 0$, a contradiction, and similarly $\rho \neq 0$ if $\overline{a}^T\check{x} \leqslant 0$, $A^T\check{x} < 0$, and $\overline{a} \neq 0$ (otherwise $0 \geqslant \overline{a}^T\check{x} = -\overline{\mu}^TA^T\hat{x} \geqslant 0$ would yield $\overline{\mu} = 0$ and $\overline{a} = 0$). Hence $\overline{G}^{-1} \geqslant 0$ by (ii) and Lemma 1.1. (v) and (vi) follow from (i), since $\overline{\mu} = -G^{-1}A^T\overline{a}$. ■

COROLLARY 3.3 [Tod79, Theorem 4]. *If $A \in \mathbb{R}^{n\times m}$ is such that $G = A^TA$ has nonpositive off-diagonal entries then $G$ is monotone $\Leftrightarrow A^T\check{x} < 0$ for some $\check{x} \Leftrightarrow R$ is monotone, where $R$ is the unique Cholesky factor of $G$ having a positive diagonal.*

*Proof.* If $A^T\check{x} < 0$, apply Lemma 3.2 inductively to get $G^{-1} \geqslant 0$ and $R^{-1} \geqslant 0$. If $G^{-1}$ exists, let $\check{x} = -AG^{-1}e$. Finally, if $R^{-1} \geqslant 0$ then $G^{-1} = R^{-1}R^{-T} \geqslant 0$. ■

A *surrogate projection method* [Kiw94] for finding a point in $\mathscr{P} = \{x\colon A^Tx \leqslant b\}$ runs as follows. At iteration $k$, given the current iterate $x^k$, $s^k = Ax^k - b \nleqslant 0$ and a control index $i_k$ such that $s^k_{i_k} > 0$ (e.g., $i_k \in \operatorname{Arg\,max}_i s^k_i$), choose $I = I^k \subset \{1{:}m\}$ such that $i_k \in I^k$ and $G_{II} = A_I^TA_I$ is monotone, pick a stepsize $t_k \in (0,2)$, and set $x^{k+1} = x^k - t_kA_I(A_I^TA_I)^{-1}s^k_I = x^k + t_k[P_{\mathscr{P}_I}(x^k) - x^k]$ (cf. Lemma 2.1). To choose $I^k$, starting with $G_{II}^{-1} \geqslant 0$ and $s^k_I \geqslant 0$ (e.g., $I = \{i_k\}$), we may repeatedly augument $I$ with any $i \in \{1{:}m\} \setminus I$ such that $A_I^\dagger a^i \leqslant 0$ and either $s^k_i > 0$, or $s^k_i = 0$ and either $s^k_I > 0$ or rank $A_{I\cup\{i\}} = |I| + 1$, detecting $\mathscr{P} = \varnothing$ if rank $A_{I\cup\{i\}} = |I|$ (since rank $A_{I\cup\{i\}} = |I| + 1$ if $\mathscr{P} \neq \varnothing$ by Lemma 3.2(iv) with $\check{x} \in \mathscr{P} - x^k$). It is natural to choose $I^k$ as large as possible, although one need not insist on maximality. Note that checking $A_I^Ta^i \leqslant 0$ as in [Tod79] instead of $A_I^\dagger a^i \leqslant 0$ (cf. Lemma 3.2(v)) may unnecessarily reject $i$ if $A_I^Ta^i \nleqslant 0$ but $A_I^\dagger a^i \leqslant 0$, whereas the cost of computing $A_I^\dagger a^i$ need not be much higher than that of $A_I^Ta^i$.

Using the Cholesky factorization $G_{II} = R^T R$, we may compute $\overline{\mu}_I = -A_I^{\dagger} a^i$ from $R^T R \overline{\mu}_I = -A_I^T a^i$ and update $R$ as in Lemma 3.2, finally finding $\overline{\lambda}_I = G_{II}^{-1} s_I^k$ from $R^T R \overline{\lambda}_I = s_I^k$. Iterative refinement may be employed to improve accuracy in the presence of rounding errors. Alternatively, one may use any stable method, e.g., the Gram–Schmidt process with reorthogonalization, to compute the "skinny" QR-factorization $A_I = QR$, where $Q$ is orthonormal. In practice detecting rank $A_{I \cup \{i\}} = |I|$ will require tolerances tuned to the factorization of $A_I$. All these aspects are treated in depth in, e.g., [Bjö90, Fle87, GMW91, GVL89].

*This research was supported by the State Committee for Scientific Research under Grant 8S50502206.*

REFERENCES

[Agm54] S. Agmon, The relaxation method for linear inequalities, *Canad. J. Math.* **6**:382–392 (1954).

[BB96] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev., in press.

[BGT81] R. B. Bland, D. Goldfarb, and M. J. Todd, The ellipsoid method: A survey, *Oper. Res.* **29**:1039–1091 (1981).

[Bjö90] Å. Björck, Least squares methods, in *Finite Difference Methods (Part 1): Solution of Equations in $R^n$*, Handbook of Numerical Analysis, Vol. I (P. G. Ciarlet and J. L. Lions, Eds.), North-Holland, Amsterdam, 1990, pp. 465–652.

[Cen81] Y. Censor, Row-action methods for huge and sparse systems and their applications, *SIAM Rev.* **23**:444–466 (1981).

[DBMS79] J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users' Guide*, SIAM Publications, Philadelphia, 1979.

[DPI85] A. R. De Pierro and A. N. Iusem, A simultaneous projections method for linear inequalities, *Linear Algebra Appl.* **64**:243–253 (1985).

[Fle87] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., Wiley, Chichester, 1987.

[GMW91] P. E. Gill, W. Murray, and M. H. Wright, *Numerical Linear Algebra and Optimization*, Addison–Wesley, Redwood City, CA, 1991.

[Gof80] J.-L. Goffin, The relaxation method for solving systems of linear inequalities, *Math. Oper. Res.* **5**:388–414 (1980).

[GoT82] D. Goldfarb and M. J. Todd, Modifications and implementation of the ellipsoid algorithm for linear programming, *Math. Programming* **23**:1–19 (1982).

[GP93] U. García-Palomares, Parallel projected aggregation methods for solving the convex feasibility problem, *SIAM J. Optim.* **3**:882–900 (1993).

[GVL89] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins Univ. Press, Baltimore, MD, 1989.

[Kiw95] K. C. Kiwiel, Block-iterative surrogate projection methods for convex feasibility problems, *Linear Algebra Appl.*, 215:225–260 (1995).

[Kiw96a] K. C. Kiwiel, The efficiency of subgradient projection methods for convex optimization, part I: General level methods, *SIAM J. Control Optim.* 34:660–676 (1996).

[Kiw96b] K. C. Kiwiel, The efficiency of subgradient projection methods for convex optimization, part II: implementations and extensions, *SIAM J. Control Optim.* 34:677–697 (1996).

[LaH74] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice–Hall, Englewood Cliffs, NJ, 1974.

[Mer62] Yu. I. Merzlyakov, On a relaxation method of solving systems of linear inequalities, *Zh. Vychisl. Mat. i Mat. Fiz.* **2**:482–487 (1962) [in Russian]; English transl. in *USSR Comput. Math. Math. Phys.* **2**:504–510 (1962).

[MoS54] T. Motzkin and I. J. Schoenberg, The relaxation method for linear inequalities, *Canad. J. Math.* **6**:393–404 (1954).

[Oko92] S. O. Oko, Surrogate methods for linear inequalities, *J. Optim. Theory Appl.* **72**:247–268 (1992).

[Ott88] N. Ottavy, Strong convergence of projection-like methods in Hilbert spaces, *J. Optim. Theory Appl.* **56**:433–461 (1988).

[Tod79] M. J. Todd, Some Remarks on the Relaxation Method for Linear Inequalities, Technical Report 419, Cornell Univ., Ithaca, NY, 1979.

[YaM92] K. Yang and K. G. Murty, New iterative methods for linear inequalities, *J. Optim. Theory Appl.* **72**:163–185 (1992).

*Received 10 June 1994; final manuscript accepted 1 August 1995*